Informan : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi
Vol.8 No.1 Januari 2023, P-ISSN:2502-3470, E-ISSN:2581-0367

Perbandingan Hasil Uji Stemming Algoritma Tala dengan Nazief Adriani dalam Dokumen Abstrak dan Berita Nasional

Natalinda Pamungkas[1*], Erika Devi Udayanti[2], Bonifacius Vicky Indriyono[3], Wildan Mahmud[4], Ery Mintorini[5], Arika Norma Wahyu Dorroty[6], Sanina Quamila Putri[7]

[1,2,3,4,6,7]*Sistem Informasi, Universitas Dian Nuswantoro, Semarang, Indonesia*

[5]*Desain Komunikasi Visual, Universitas Dian Nuswantoro, Semarang, Indonesia*

[1]natalinda.pamungkas@dsn.dinus.ac.id (*)

[2,3,4,5][erikadevi, ery.mintorini, wildan.mahmud, bonifacius.vicky.indriyono]@dsn.dinus.ac.id

[6,7][norma.wahyu08, saninap01]@gmail.com

Diterima: 13-12-2022; Diterima: 09-01-2023; Diterbitkan: 28-01-2023

Abstrak—Keberadaan informasi tidak dapat dipungkiri lagi dibutuhkan oleh banyak orang. Pernyataan ini menggambarkan semakin pentingnya informasi dan peningkatan kebutuhan akan akses terhadap dokumen dan literatur yang relevan. Isi informasi yang diperoleh dari dokumen-dokumen tersebut kemudian dipilah-pilah agar lebih mudah dipahami maknanya. Proses penyortiran ini disebut stemming. Stemming merupakan proses yang banyak diterapkan dalam pencarian kata dasar. Memisahkan kata-kata yang tidak bermakna dapat membuat informasi menjadi lebih jelas. Perlu diperhatikan algoritma stemming yang sesuai dengan bahasa yang digunakan. Banyak algoritma stemming yang dapat digunakan untuk melakukan proses pencarian kata dasar ini. Beberapa di antaranya adalah algoritma Tala dan Nazief Adriani. Kedua algoritma tersebut memiliki perbedaan dalam proses kerjanya. Algoritma Tala mengadopsi algoritma Porter berbasis aturan, sedangkan algoritma Nazief & Adriani bekerja berdasarkan kamus. Kedua algoritma tersebut memiliki keunggulan masing-masing dalam hal akurasi dan kecepatan. Oleh karena itu, pada penelitian ini akan dilakukan analisis dengan membandingkan kinerja kedua algoritma pada proses stemming teks bahasa Indonesia. Proses uji cobanya menggunakan beberapa sumber data yang berbeda untuk mengukur kecepatan dan keakuratan setiap algoritma. Sumber data yang digunakan dalam penelitian ini antara lain abstrak laporan skripsi atau tugas akhir mahasiswa sebanyak 30 mahasiswa dan informasi dari berita online sebanyak 200. Dari hasil pengujian yang telah dilakukan dapat disimpulkan bahwa algoritma stemming Tala mempunyai kinerja yang baik. tingkat akurasi yang lebih rendah dibandingkan Nazief Adriani. Algoritma Tala hanya memiliki rata-rata akurasi sebesar 65,29%, sedangkan Nazief Adriani memiliki akurasi sebesar 78,47%. Dari segi kecepatan, algoritma Tala memiliki kecepatan lebih baik dibandingkan Nazief Adriani sebesar 32,19 detik dan Nazief & Adriani sebesar 65,2 detik.

Kata kunci— Stemming, Algoritma Nazief Adriani, Algoritma Tala, Dokumen Abstrak, Berita Nasional

saya.sayaPENDAHULUAN

Kebutuhan akan informasi di era teknologi saat ini sangat dibutuhkan oleh para penggunanya. Teknologi ini digunakan dalam pencarian dokumen teks, khususnya di internet, untuk mendapatkan informasi. Permasalahan yang sering muncul dalam mendapatkan informasi adalah mencari informasi yang sesuai dengan kebutuhan. Informasi yang tepat dapat diperoleh dengan melakukan pemisahan kata pada dokumen teks. Salah satu cara memisahkan kata adalah dengan mendapatkan informasi menggunakan proses stemming. Stemming merupakan salah satu proses transformasi kata dalam teks menjadi kata dasar atau menghilangkan imbuhan kata [1]. Algoritme stemming untuk satu bahasa akan berbeda dengan bahasa lainnya. Algoritma Nazief Adriani dan Tala paling banyak digunakan untuk stemming dokumen dalam bahasa Indonesia. Algoritma Nazief Adriani merupakan algoritma stemming yang prinsip kerjanya menggunakan kamus, sedangkan algoritma Tala mengadopsi algoritma Porter dan mempunyai prinsip kerja rulebased. Proses stemming dalam teks bahasa Indonesia mempunyai beragam imbuhan yang harus dihilangkan untuk mendapatkan akar kata suatu kata [2]. Stemming juga dapat digunakan dalam proses pembelajaran bahasa Indonesia mengenai kata dasar [3]. Nopiyanti [4] meneliti tentang pembentukan aplikasi stemming yang digunakan untuk mencari kata dasar sesuai Kamus Besar Bahasa Indonesia (KBBI). Dengan menggunakan algoritma Porter pada proses stemming 20 dokumen, masih dapat dikembangkan dengan melakukan pengujian terhadap lebih banyak dokumen dan dataset berita. Penelitian yang dilakukan [1] menganalisis tahapan stemming menggunakan algoritma Porter untuk dokumen berbahasa Indonesia yang menggunakan dokumen berekstensi/format .txt dan dataset kamus yang tidak lengkap. Dari penelitian ini masih dimungkinkan untuk dikembangkan pengujian apakah dokumen yang digunakan dapat berekstensi .pdf/.docx untuk mengetahui keakuratan yang ingin diperoleh. Dengan kamus yang lebih lengkap, hasil tes yang akurat dapat dikembangkan. Utomo [5] meneliti dengan menggunakan algoritma pada korpus (abstrak). Penelitian ini didasarkan pada aturan-aturan yang terdapat pada algoritma Tala. Penelitian ini masih memiliki tingkat kesalahan yang tinggi pada proses stemmingnya, sehingga masih dapat dikembangkan. Penelitian peningkatan kemampuan batang dengan menambahkan dua tingkat morfologi dilakukan oleh [6] memperoleh akurasi yang cukup tinggi. Pada penelitian ini hanya menggunakan sepuluh dokumen, pengembangan masih dapat dilakukan secara komparatif jika dilakukan pengujian terhadap lebih banyak dataset berita dan dokumen. Dari penelitian yang dilakukan [2], membandingkan hasil algoritma Porter dan Nazief & Adriani masih menggunakan dokumen saja untuk pengujiannya. Ada kemungkinan yang dapat dilakukan untuk mendapatkan akurasi pengujian dari beberapa sampel dataset berita untuk menjadi pengujian yang dapat dipertimbangkan.

Beberapa penelitian sebelumnya yang membahas topik kinerja algoritma stemming. Jurnal ini [3] menjelaskan prosesnya

DOI : http://dx.doi.org/10.25139/inform.v8i1.5569

pembuatan perangkat lunak stemming untuk mencari kumpulan kata dasar yang sesuai dengan yang tersimpan dalam Kamus Besar Bahasa Indonesia (KBBI) dari dokumen teks bahasa Indonesia dengan menggunakan porter stemmer. Peneliti [7] menjelaskan proses stemmer dalam menentukan klasifikasi jenis buku menggunakan algoritma Porter-Stemmer. Penelitian [7] membahas penggunaan algoritma Porter untuk mengklasifikasikan jenis buku secara otomatis berdasarkan aturan dasar pencarian kata. Hasil penelitian menunjukkan bahwa algoritma tersebut efektif dalam mengklasifikasikan jenis buku secara akurat. Namun, penelitian [4] menyebutkan bahwa algoritma berbasis aturan dapat memiliki tingkat kesalahan yang tinggi, yang dapat berdampak negatif pada keakuratan hasil akhir. Hasil penelitian menyimpulkan [8] proses pencarian diperoleh rata-rata perhitungan waktu respon aplikasi sebesar 5,66 detik, dan hasil recall pencarian dokumen diperoleh rata-rata sebesar 83%. Penelitian [9] pada algoritma Stemming untuk tenses berbeda untuk menyempurnakan kamus bahasa Persia menjelaskan algoritma stemming berbasis aturan, tidak menggunakan kamus. Penelitian ini menyimpulkan [9] bahwa algoritma yang digunakan memiliki akurasi untuk kata kerja biasa dalam bentuk simple/past, continuous dan perfect tense namun dibatasi hanya pada kamus dengan kata kerja beraturan dan terbatas pada pengujian 50 kata dari 465 kata yang ada di kamus. Algoritma Adriani & Nazief dan Algoritma Kemiripan dapat digunakan untuk memeriksa judul dan abstraksi skripsi [10] dan apakah judul dengan tema tersebut sudah diajukan. Merupakan stemming yang berfungsi untuk mengumpulkan indeks judul dan abstraksi skripsi sebagai database sehingga dapat diperiksa menggunakan algoritma kemiripan. Penelitian [11] terkait deteksi kemiripan teks menyimpulkan bahwa penerapan metode stemming Nazief Adriani pada algoritma Rabin-Karp sangat mempengaruhi tingkat persentase kemiripan teks sehingga memudahkan dalam mendeteksi kemiripan teks.

Berdasarkan literatur penelitian terdahulu mengenai proses stemming, penelitian ini akan membandingkan hasil uji stemming Tala dengan Nazief Adriani pada dokumen abstrak dan berita nasional sebagai sumber data pengujian.

II. METODOLOGI PENELITIAN

Metodologi penelitian adalah suatu metode yang dapat dipelajari secara ilmiah dan digunakan oleh peneliti untuk memperoleh data guna menunjang kegiatan penelitian atau untuk tujuan lain [12].

A.Metode Penelitian

Penelitian ini dilakukan dengan menggunakan metode eksperimen. Menurut [13], penelitian eksperimen merupakan metode penelitian yang menggambarkan hubungan sebab akibat, sehingga metode ini dapat dikatakan sebagai metode penelitian sebab akibat. Secara khusus metode penelitian yang dilakukan dapat dijelaskan pada Gambar 1, maka dapat dijelaskan tahapan penelitiannya adalah sebagai berikut:

1. Studi Sastra: Tahap pertama meliputi pengumpulan beberapa referensi pendukung penulisan topik penelitian. Sumber-sumber ini dapat mencakup jurnal akademis, buku, prosiding konferensi, dan artikel lainnya.

2. Pengumpulan Data: Tahap penelitian selanjutnya adalah pengumpulan data. Data yang digunakan untuk melaksanakan penelitian ini berasal dari intisari laporan skripsi atau tugas akhir mahasiswa, informasi dari berita online, dan Twitter nasional Indonesia.

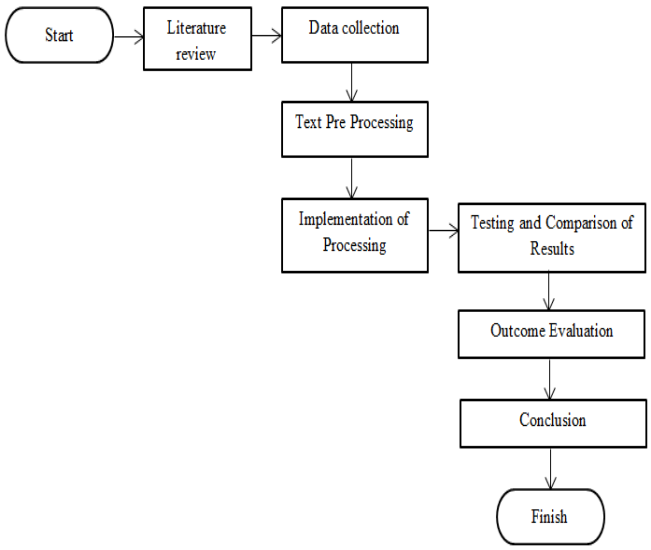
3. Pra-pemrosesan Teks: Proses ini membantu menjadikan data teks lebih konsisten dan terstruktur untuk analisis dan menghilangkan informasi yang tidak relevan atau berlebihan yang mungkin memengaruhi hasil. Kegiatan ini mencakup (i) pelipatan huruf besar-kecil: mengubah semua karakter atau huruf kapital menjadi huruf kecil, (ii) tokenisasi: memecah string atau kalimat menjadi kata-kata individual, dan (iii) menghentikan penghapusan kata: menghilangkan kata-kata yang tidak berarti dari daftar token yang dihasilkan selama tokenisasi .

4. Penerapan Teknik Pengolahan: Tahap selanjutnya adalah menentukan teknik atau metode yang digunakan untuk melakukan proses pra-pemrosesan teks. Metode yang digunakan adalah Nazief Adriani dan Tala.

5. Menguji dan Membandingkan Hasil: Setelah metode ditentukan, dilakukan uji stemming dengan menggunakan kedua algoritma tersebut. Pengujian dilakukan untuk melihat hasil dan membandingkan hasil tersebut baik dari segi akurasi maupun kecepatan.

6. Evaluasi Hasil: Setelah dilakukan pengujian dan perbandingan hasilnya, dilakukan evaluasi. Evaluasi berkaitan dengan keakuratan dan kecepatan kedua metode yang digunakan untuk melakukan proses stemming.

7. Kesimpulan: Tahap terakhir adalah menyimpulkan evaluasi yang telah dilakukan. Kesimpulan ini terkait dengan perbandingan algoritma dengan akurasi dan kecepatan tinggi dalam proses stemming.



Gambar 1. Metode Penelitian

B. Metode Pengumpulan Data

Metode pengumpulan data diartikan sebagai cara yang digunakan oleh para ilmuwan untuk memperoleh data guna menunjang kegiatan penelitiannya. Metode ini mengarah pada asal data yang ingin diperoleh. Sumber utamanya adalah wawancara dan observasi atau sumber pendukung berupa studi literatur melalui

DOI : http://dx.doi.org/10.25139/inform.v8i1.5569

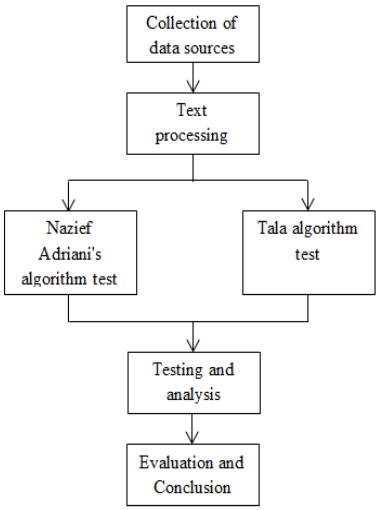
buku, jurnal, dan prosiding [14]. Berdasarkan pengertian teknik pengumpulan data, data yang diperoleh dalam kegiatan ini berasal dari sumber pendukung yaitu dari beberapa skripsi dan tugas akhir mahasiswa, berita online, Twitter, buku, jurnal, dan prosiding yang membahas tentang proses stemming teks bahasa Indonesia. Data yang diperoleh akan dijadikan acuan untuk menganalisis keakuratan dan kecepatan proses stemming dari algoritma Nazief Adriani dan Tala.

C. Membendung

Algoritme stemming memetakan varian morfologi yang berbeda dari sekumpulan kata menjadi kata dasar. Algoritma stemming ini banyak digunakan dalam linguistik komputasi dan pengambilan informasi [15]. Stemming juga dapat diartikan menghilangkan imbuhan, prefiks, sufiks, dan preposisi. Hasil dari proses penghapusan stopword sehingga istilah dapat menjadi bentuk kata dasar [6], sedangkan menurut [16], stemming adalah proses pencarian kata dasar yang digunakan untuk mendefinisikan ciri-ciri dalam teks. Gambar 2 menunjukkan apa sebenarnya yang dimaksud dengan proses stemming.

Gambar 2. Contoh Stemming

Penelitian ini bertujuan untuk menganalisis kinerja algoritma Nazief Adriani dengan Tala untuk proses textstemming bahasa Indonesia. Hasil yang diharapkan adalah mengetahui reliabilitas kedua algoritma yang digunakan dari segi akurasi dan kecepatan pemrosesan. Gambar 3 secara eksplisit menunjukkan pola pikir proses pengujian stemming. Sumber data yang digunakan dalam penelitian ini berasal dari situs berita online Tribunnews.com dan kutipan laporan skripsi mahasiswa.



Gambar 3. Alur Pemikiran Proses Stemming

D. Algoritma Nazief Adriani

Algoritma Nazief Adriani merupakan salah satu dari beberapa algoritma stemming yang ditemukan oleh Bobby Nazief dan Mirna Adriani.

Menurut [2], beberapa tahapan atau langkah penyelesaian menggunakan algoritma Nazief Adriani antara lain:

1. Carilah kata-kata yang berasal dari kamus yang telah disiapkan. Jika kata-kata tersebut ada di kamus, disimpulkan bahwa kata tersebut merupakan kata dasar, dan prosesnya terhenti.
2. Membuang Sufiks seperti "*-lah*", "*-kah*", "*-ku*", "*-mu*", atau "*-nya*". Jika sufiksnya berupa partikel seperti "*-lah*",'*-kah*","*- tah*" atau "*-permainan kata-kata*" lalu ulangi langkah untuk menghilangkan kata ganti posesif seperti "*-ku*", "*-mu*", atau "*-nya*", ketika ditemukan.
3. Menghilangkan Akhiran Derivasi ("-i", "-an", atau "*-ka*N"). Jika akhiran ditemukan di kamus, proses akan berhenti, tetapi jika tidak ditemukan, proses akan dilanjutkan ke langkah 3.1.
   3.1 Jika akhiran "-an" dihapus dan kata berakhiran "-k", maka akhiran tersebut juga dihapus. Jika kata tersebut ditemukan dalam kamus, algoritma berhenti. Jika tidak ditemukan, maka lakukan langkah 3.2.
   3.2 Mengembalikan sufiks yang dihapus seperti "-i", "-an", atau "*-kan*"
4. Hapus Awalan Derivasi. Jika pada langkah 3 diatas terdapat sufiks yang dihapus maka proses dilanjutkan ke langkah 4.1, namun jika tidak ada sufiks yang dihapus maka proses dilanjutkan ke langkah 4.2.
   4.1 Melakukan pemeriksaan terhadap Tabel yang mengandung kombinasi awalan dan akhiran yang tidak boleh digunakan apabila terdapat pada Tabel. Prosesnya berhenti; sebaliknya jika tidak ditemukan, lanjutkan ke langkah 4.2.
   4.2 Tentukan jenis awalan dan kemudian hapus awalan tersebut. Apabila akar kata belum ditemukan maka proses dilanjutkan pada langkah ke 5, namun sebaliknya jika akar kata ditemukan maka proses dihentikan dengan catatan posisi awalan kedua sama dengan awalan pertama.
5. Lakukan perekaman langkah
6. Apabila seluruh tahapan telah selesai namun belum membuahkan hasil, maka kata awal dapat disimpulkan sebagai kata dasar. Proses selesai.

Langkah-langkah dalam menentukan jenis awalan pada algoritma Nazief Adriani adalah sebagai berikut :

1. Awalan "*di-*", "*ke-*", atau "*se-*"memiliki tipe awalan"*di-*", "*ke-*", atau "*se-*".
2. Diperlukan proses tambahan untuk menentukan jenis awalan jika awalan tersebut dikenal dengan nama "*te-*", "*Saya-*", "*menjadi-* ", atau "*pe-*".
3. Proses akan berhenti jika dua karakter pertama tidak "*di-*","*ke-*","*se-*","*te-*","*menjadi-*","*Saya-*",*atau* "*pe-*".
4. Ketikkan awalan "none" lalu proses berhenti. Namun jika jenis prefiksnya bukan "none", maka penentuan prefiksnya dapat dilihat pada Tabel II. Awalan dapat dihapus jika ditemukan.

Berikut disajikan daftar kombinasi prefiks dan sufiks yang tidak boleh digunakan, cara menentukan jenis prefiks "te" dan kelompok prefiks berdasarkan jenis prefiksnya seperti terlihat pada Tabel I, II, dan III.

DOI : http://dx.doi.org/10.25139/inform.v8i1.5569

TMAMPUSAYA
TIDAK DIIZINKAN KOMBINASI PREFIGURE DAN KECUKUPAN

TIDAK	Awalan	Akhir yang tidak pantas
1	menjadi-	- Saya
2	di-	- sebuah
3	ke-	- aku,-kan
4	Saya-	- sebuah
5	se-	- aku, -kan

TMAMPUII
PENENTUAN JENIS"*TE*"

Karakter Berikut				Awalan
Tetapkan 1	Tetapkan 2	Atur 3	Tetapkan 4	Jenis
- R-	"-R-"	-	-	tidak ada
- R-	Vokal	-	-	*ter-luluh*
- R-	bukan (vokal atau "-R-")	"-er-"	vokal	*ter*
- R-	bukan (vokal atau "-R-")	"-er-"	bukan vokal	*ter-*
- R-	bukan (vokal atau "-R-")	bukan "-er-"	-	*ter*
tidak (vokal atau "-r-")	"-er-"	Vokal	-	tidak ada
tidak (vokal atau "-r-")	"-er-"	bukan vokal	-	*te*

TMAMPUAKU AKU AKU
KELOMPOK PREFIX MENURUT JENIS PREFIX

Awalan	Awalan yang Dihapus
di-	*di-*
ke-	*ke-*
se-	*se-*
te-	*te-*
ter-	*ter-*
ter-luluh	*Ter*

E.Algoritma Tala

Algoritma Tala melakukan pemrosesan dari awalan, akhiran, dan kombinasi awalan dan akhiran pada kata turunan. Stemming Tuning merupakan algoritma steeming yang diadopsi dari algoritma stemming khusus bahasa Inggris yaitu Porter's stemming. Kapal uap Tala Indonesia mempunyai struktur pembentukan kata bahasa Indonesia sebagai berikut [17] : **[awalan-1] + [awalan-2] + dasar + [akhiran] + [milik] + [membawa],**dimana masing-masing bagian tersebut digabungkan dengan akar kata sehingga membentuk sebuah kata yang mempunyai pahala. Algoritma Tala memiliki tiga langkah awal dan dua langkah opsional, sebagai berikut:

1. Lakukan penghilangan partikel
2. Hilangkan Kata Ganti Posesif
3. Awalan pertama dihilangkan. Jika tidak ditemukan proses lanjutkan ke langkah 4.1, namun jika ada lakukan pencarian dan proses akan dilanjutkan ke langkah 4.2
4. Tahapan langkah 4
   - 4.1. Hapus awalan kedua dan lanjutkan ke langkah 5.1
   - 4.2. Hapus akhiran. Jika tidak ditemukan maka kata yang dicari disimpulkan sebagai kata dasar, namun jika ditemukan lanjutkan ke langkah 5.2.
5. Tahapan langkah 5
   - 5.1. Menghilangkan akhiran dan kata akhir disimpulkan sebagai kata dasar.
   - 5.2. Awalan kedua dihilangkan, dan kata terakhir disimpulkan sebagai kata dasar.

Algoritma Tala memiliki lima kategori aturan imbuhan yang ditunjukkan pada Tabel IV hingga Tabel VIII.

TMAMPUIV
ATURAN PARTIKEL INFLEKSI

Akhiran	Penggantian	Ukuran Kondisi	Tambahan Kondisi	Contoh
- kah	BATAL	2	BATAL	*bukukah*
- lah	BATAL	2	BATAL	*pergilah*
- permainan kata-kata	BATAL	2	BATAL	*bukupun*

TMAMPUV
ATURAN KATA GANTI POSSESIF INFLEKSI

Akhiran	Penggantian	Ukuran Kondisi	Tambahan Kondisi	Contoh
-ku	BATAL	2	BATAL	*bukuku*
- mu	BATAL	2	BATAL	*bukumu*
-nya	BATAL	2	BATAL	*bukunya*

TMAMPUVI
ATURAN UNTUK PERTAMA-Awalan derivasional pesanan

Akhiran	Penggantian	Ukuran Kondisi	Tambahan Kondisi	Contoh	
meng-	BATAL	2	BATAL	*mengukur*	*ukur*
meny-	S	2	V...*	*menyapu*	*sapu*
laki-laki-	BATAL	2	BATAL	*menduga*	*dugaan*
mem-	P	2	V ...	*memaksa*	*paksa*
mem-	BATAL	2	BATAL	*membaca*	*Baca*
Saya-	BATAL	2	BATAL	*merusak*	*rusak*
peng-	BATAL	2	BATAL	*pengukur*	*ukur*
sen-	S	2	V ...	*penyapu*	*sapu*
pena-	BATAL	2	BATAL	*penduga*	*dugaan*
pem-	P	2	V ...	*pemaksa*	*paksa*
pem-	BATAL	2	BATAL	*Pembaca*	*Baca*
di-	BATAL	2	BATAL	*diukur*	*ukur*
ter-	BATAL	2	BATAL	*tersapu*	*sapu*
ke-	BATAL	2	BATAL	*kekasih*	*terima kasih*

TMAMPUVII
ATURAN KEDUA-Awalan derivasional pesanan

Akhiran	Penggantian	Ukuran Kondisi	Tambahan Kondisi	Contoh	
ber-	BATAL	2	BATAL	*berlari*	*lari*
Bel-	BATAL	2	Terbuka sedikit	*belajar*	*terbuka sedikit*
menjadi-	BATAL	2	oke	*bekerja*	*kerja*
per-	BATAL	2	BATAL	*jelas*	*jelas*
Pel-	BATAL	2	Terbuka sedikit	*pelajar*	*terbuka sedikit*
pe-	BATAL	2	BATAL	*pekerja*	*kerja*

TMAMPUVIII
ATURAN SUFFIX DERIVASI

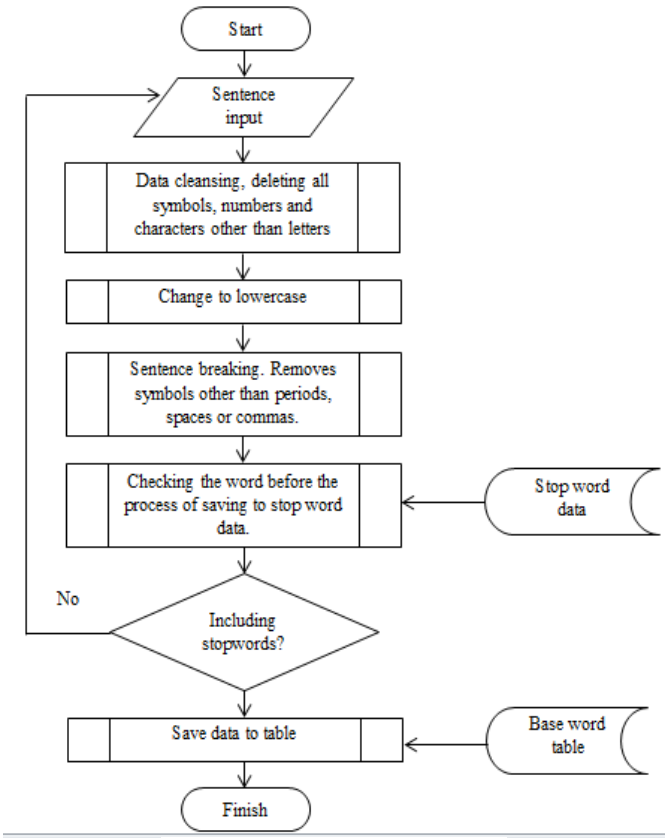
Akhiran	Ganti-semen	Ukuran kondisi	Tambahan kondisi	Contoh	
- kan	batal	2	awalan tidak seorang anggota {*ke, peng*}	*tarikkan*	*tarik*
				mengambilkan	*ambil*
			awalan awalan	*makanan*	*makan*
- sebuah	batal	2	bukan sebuah anggota {*di, meng, ter*}	.	*janji*
- Saya	batal	2	awalan tidak	*tandai*	*tanda*

DOI : http://dx.doi.org/10.25139/inform.v8i1.5569

Akhiran	Ganti-semen	Ukuran kondisi	Tambahan kondisi	Contoh	
			seorang anggota {*ber, ke, peng*}	*ternyata*	*dapat*

F. Pra-pemrosesan Teks

Pada tahap pra-pemrosesan teks, tahap pertama adalah pembersihan data yang telah dikumpulkan, kemudian tahap pelipatan kasus, tokenisasi, dan penghapusan stopword. Proses pembersihan adalah proses menghilangkan karakter, huruf, dan simbol yang berada di luar huruf abjad pada teks. Huruf atau karakter yang dihapus dapat berupa hak cipta, akhiran berita, dan simbol hak cipta. Kemudian setelah dibersihkan dilakukan proses selanjutnya yaitu pelipatan case. Case Folding adalah proses menyamakan seluruh huruf menjadi huruf kecil. Flowchart pengolahan teks ada pada Gambar 4.



Gambar 4. Diagram Alir Teks Pra-pemrosesan

Dari diagram alir teks pra-pemrosesan di atas dapat diberikan penjelasan sebagai berikut:

1. Masukkan Kalimat atau Teks: Teks dapat berasal dari intisari laporan tugas akhir atau skripsi mahasiswa dan teks berita yang diperoleh dari situs Tribunnews.com

2. Proses pembersihan. Teks atau kalimat yang dimasukkan akan diperiksa apakah ada karakter atau simbol lain. Jika ada maka karakter atau simbol tersebut akan terhapus.

*3. Ubah ke Huruf Kecil:*Setelah dibersihkan dan teks dianggap bersih. Kapitalisasi huruf akan disamakan. Yakni semua huruf akan diubah menjadi huruf kecil. Hal ini bertujuan untuk memudahkan proses stemming dan penghapusan stopword.

*4. Pemutusan Kalimat:*Proses pemecahan kalimat biasa disebut dengan stemming, yaitu proses yang dilakukan untuk memecah kalimat menjadi kata-kata. Pemisahan ini didasarkan pada spasi, titik, atau koma. Selain tanda tersebut, karakter lain akan dihapus.

5. Pemeriksaan Kata: Tahap ini untuk mencatat apakah kata hasil proses stemming akan diabaikan pada proses penghapusan Stopword.

6. Simpan Kata-katanya: Diabaikan pada proses penghapusan Stopword ke dalam Tabel kata dasar

Proses stemming dapat dimulai setelah menyelesaikan dua tahap yang dijelaskan di atas. Proses stemming digunakan untuk memisahkan kalimat menjadi satu kata, dan akan dilakukan proses pengecekan untuk mengetahui apakah suatu kata stemming menghasilkan kata yang akan diabaikan pada proses penghapusan stopword. Stemming adalah proses yang digunakan untuk memisahkan kalimat menjadi kata-kata individual. Hasil penelitian Tala dijadikan data stopword pada Tabel IX.

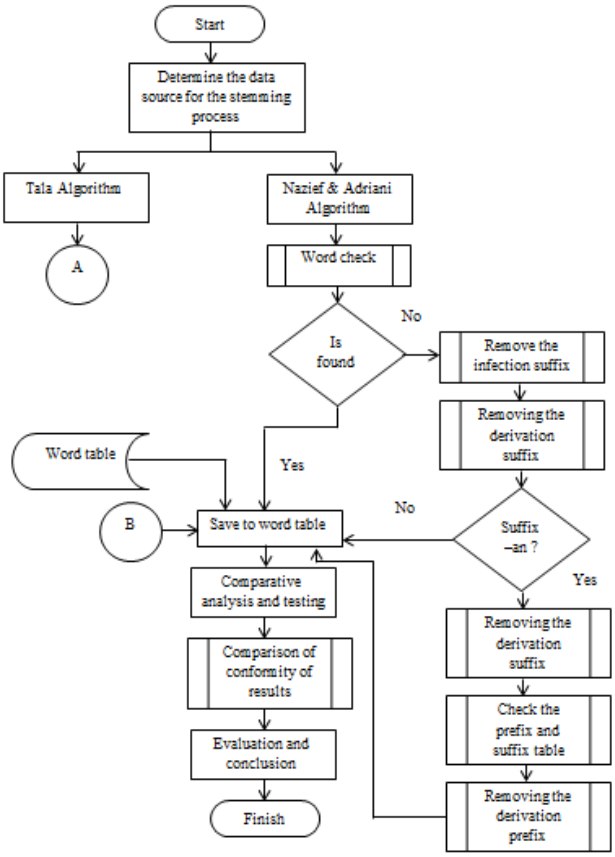
TMAMPUIX
CONTOH DATA STOPWORD(TALA2013)

ada	*amatlah*	*atas*
adalah	*anda*	*atau*
adanya	*andalah*	*orkah*
adapun	*antar*	*ataupun*
agak	*antara*	*awal*
agaknya	*diantaranya*	*awalnya*
agar-agar	*apa*	*macam-macam*
akan	*apaan*	*bagaikan*
akankah	*jika*	*Bagaimana bagaimana*
akhir	*apakah*	*bagaimanakah*
akhiri	*apalagi*	*Namun*
akhirnya	*apatah*	*bagi*
aku	*artinya*	*bagian*
akulah	*asal*	*...*
amat	*Asalkan*	

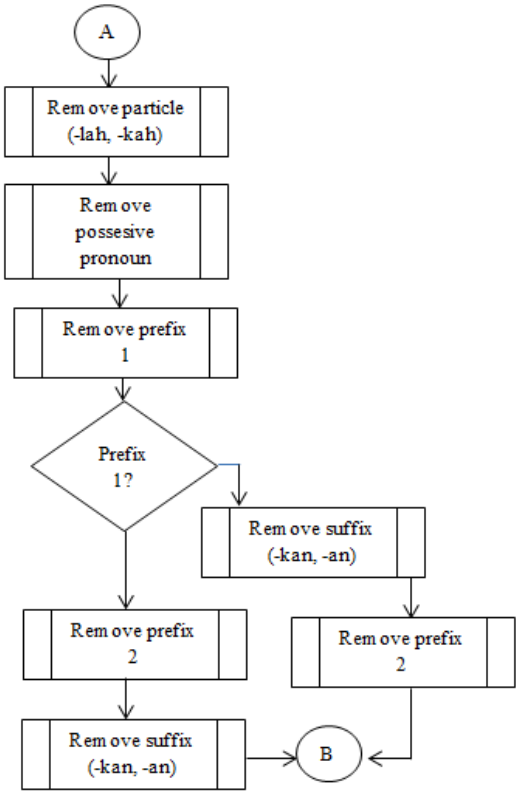
A.*Proses Pembendungan*

Stemming adalah prosedur yang merupakan bagian dari sistem pengambilan informasi (IR, yang merupakan singkatan dari "Information Retrieval"), dan tujuan dari proses ini adalah untuk mengubah kata-kata dalam teks menjadi bentuk paling mendasar dengan menerapkan seperangkat aturan. . Misalnya, "sama" ditampilkan sebagai akar kata dari "membawa", "membawa", dan "membawa", ketika kata-kata ini dikenai stemming. Menurut penulis [16], stemming adalah proses menemukan kata dasar, yang kemudian digunakan pada langkah selanjutnya menentukan ciri-ciri dalam teks. Gambar 5 dan 6 menyajikan representasi diagram alir proses stemming.

DOI : http://dx.doi.org/10.25139/inform.v8i1.5569



Gambar 5. Flowchart Stemming 1



Gambar 6. Flowchart Stemming 2

A. Uji Pengumpulan Data

Data yang digunakan untuk pengujian pada skenario ini bersumber dari website berita Tribunnews.com dan abstrak dokumen skripsi mahasiswa. Data-data tersebut akan disimpan pada database tabel X dan XI untuk diproses dan dianalisis lebih lanjut.

TMAMPUX
CONTOH BERITA DALAM DATABASETMAMPU

Id Berita	01
Nm_filenews	01.dok
URL	*https://www.tribunnews.com/nasional/2022/12/06/genap-45-tahun-bpjs-ketenagakerjaan-satukan-semangatsejahterakan-pekerja*
Judul_berita	*Genap 45 Tahun, BPJS Ketenagakerjaan Satukan Semangat Sejahterakan Pekerja*
Konten berita	*Genap memasuki usia 45 tahun, BPJS Ketenagakerjaan berikrar ùntuk terus berkembang dan bergerak maju, menjaga integritas serta menyatukan semangat mensejahterakan seluruh pekerja Indonesia. Anggoro Eko Cahyo dalam keterangannya mengatakan, berkomitmen menyatukan semangat yang datang dari seluruh insan BPJS Ketenagakerjaan dan juga dari pemangku kepentingan terdekat seperti penerbit, pengusaha hingga serikat pekerja/buruh, untuk meningkatkan kesejahteraan pekerja Indonesia. "Hari ini kami genap berusia 45 tahun, sebuah usia yang sudah bisa disebut matang, kami berikrar untuk terus memperluas cakupan program perlindungan jaminan sosial ketenagakerjaan untuk seluruh pekerja, terutama saat ini untuk pekerja informal atau pekerja bukan penerima upah, dan juga kami akan terus meningkatkan kualitas pelayanan, sehingga peserta akan semakin merasakan manfaat hadirnya BPJS Ketenagakerjaan," ucap Anggoro. Pemahaman, yang dilakukan BPJS Ketenagakerjaan saat ini adalah mengoptimalkan strategi ekstensifikasi, intensifikasi dan retensi. Memanfaatkan peluang kerja sama dengan Kementerian/Lembaga dan Pemerintah Daerah, business to business, serta pemanfaatan mesin PERISAI, dan karena target peserta adalah BPU, kampanye Kerja Keras Bebas Cemas akan digunakan untuk melindungi sebanyak-banyaknya pekerja. "Saat ini pencapaian kepesertaan aktif kami adalah sebesar 36 juta tenaga kerja atau meningkat 6 juta dari tahun sebelumnya. Angka peningkatan ini merupakan rekor tertinggi selama BPJS Ketenagakerjaan berdiri, dan target sampai dengan tahun 2026 adalah 70 juta tenaga kerja," ungkapnya. Selama tahun 2022, kinerja pelayanan BPJS Ketenagakerjaan juga terus meningkat. Komitmen perubahan pola pikir ke arah berorientasi pelanggan telah membawa perubahan terhadap kualitas manfaat dan layanan yang terasa semakin dekat dengan peserta. Tercatat tingkat kesuksesan Jaminan Hari Tua (JHT) tahun ini telah mencapai 99.58 persen, dengan rata-rata SLA masa tunggu JHT via Online atau video call kurang dari 3 hari, serta rata-rata proses klaim JHT via Jamsostek Mobile (JMO) kurang dari 15 menit. Utilisasi kanal klaim melalui aplikasi JMO juga tercatat di angka 25%, lebih tinggi dari kanal Kantor Cabang sebesar 15%, namun masih dibawah utilisasi kanal Online (video call) sebesar 60%.*

DOI : http://dx.doi.org/10.25139/inform.v8i1.5569

TAMPUXSAYA
CONTOH ABSTRAK DOKUMEN TESIS

Kddok	D01
Nm_dok	Faiz-skripsi.doc
Abstrak Isi	Penyalaan pada sepeda motor yang sekaligus juga berfungsi sebagai pengamanan sepeda motor harus dirancang dan dibuat seaman mungkin untuk menghindari hilangnya kendaraan. Peralatan yang dirancang ini untuk menghidupkan dan mematikan sepeda motor dengan sistem operasi android melalui jaringan Hotspot/Wi-Fi. Input pada sistem operasi android dilakukan pada web browser dan asisten google, lalu data disimpan dan dikontrol menggunakan mikrokontroler ESP32. Dari hasil pengujian diperoleh bahwa jarak maksimum yang dapat dicapai Wi-Fi antara android dengan mikrokontroler ESP32 yang berada pada sepeda motor untuk mengoperasikan mesin sepeda motor kurang-lebih sekitar 16 meter. Sistem ini juga membuat penyalaan mesin sepeda motor menjadi penyalaan pintar karena dalam penyalaan sepeda motor pengemudi dapat menggunakan perintah suara menggunakan asisten google. Pengamanan sistem ini tidak terbatas pada pengamanan saat sepeda motor stasioner atau pada saat sepeda motor berada diparkiran, pengamanan ini juga bisa berada pada saat kendaraan digunakan, dan ketika sepeda motor ditinggalkan pengemudi pada saat menyala, maka sepeda motor akan otomatis mati jika mikrokontroler ESP32 berada di luar jangkauan Wi-Fi/Hotspot yang berasal dari perangkat android.
	Kata Kunci : Sepeda Motor, Android, Mikrokontroler ESP32, Asisten Google

B. Tes Akurasi Dan Kecepatan

Tentukan tingkat ketelitian yang diperlukan. Hal ini dapat dilakukan dengan menghitung kata yang ditemukan dan kemudian membagi hasilnya dengan jumlah total kata yang terdapat dalam satu kumpulan data. Tampilan perbandingan hasil pengujian menggunakan kedua pendekatan tersebut dibuat untuk mengetahui tingkat presisi yang dicapai oleh setiap langkah pemrosesan setelah proses stemming. Informasi yang akan dikumpulkan akan disajikan dalam format berikut: jumlah kata, kata benar, kata salah, keakuratan, dan waktu [2].

Diperlukan data pengujian dalam jumlah besar untuk mendapatkan hasil akurasi yang benar. Tabel XII menunjukkan contoh hasil pengujian teks yang telah di-stem dengan menggunakan salah satu dari dua metode yang akan diuji dalam penelitian ini.

TAMPUXI
CONTOH HASIL UJI DENGAN SATU TEKNIK

TIDAK	Kode	Total	BENAR	PALSU	BENAR %	PALSU %	Waktu
1	012	254	180	68	70,8	26,77	65,70
2	018	110	85	30	77,2	27,27	20,99
3	019	170	155	15	91,1	8,82	40,00
4	021	127	108	15	85,0	11,81	20,20
5	023	150	90	68	60	45,33	40,14
…	…	…	…	…	…	…	…
500	026	70	63	4	90	5.714	8.054
	Total	58.959	46,4	12,5	39, 6	12,02	58,95
	Rata-rata	115.430	91,7	24,7	78,5	21,53	23,75

Data pada Tabel XII merupakan contoh hasil pengujian dengan sumber data berita *tribunnews.com* yang memiliki 500 berita cerita dengan total 58.959 kata. Rata-rata yang diperoleh adalah sekitar 116 kata pada dataset berita dengan rata-rata akurasi dataset berita sebesar 78.46%.

C. Analisis Hasil Uji

Pada penelitian ini, sumber data yang digunakan untuk proses pengujian akurasi menggunakan metode Nazief Andriani dan Tala berasal dari data berita di website Tribunnews.com dan abstrak laporan skripsi mahasiswa. Variabel pengukuran nilai akurasi didasarkan pada penelitian sebelumnya yang dilakukan oleh [2]. Pola pengukurannya diperoleh dari jumlah kata yang ditemukan dibandingkan dengan total kata yang ada pada dataset. Proses pengujian dilakukan dengan total 4 skenario untuk setiap metode stemming yang digunakan. Pengujian dilakukan untuk mendapatkan nilai akurasi dan kecepatan. Data skenario pengujian terdapat pada Tabel XIII.

TAMPUXII
DATA SKENARIO UJI

Skenario	Sumber data	Dokumen	Algoritma
X1	Teks berita dari situs Tribunnews.com	506	Nazief & Adriani
X2	Teks berita dari situs Tribunnews.com	506	Tala
Y1	Teks laporan tesis abstrak	50	Nazief & Adriani
Y2	Teks laporan tesis abstrak	50	Tala

Berdasarkan data pada Tabel XII maka akan dilakukan skenario pengujian pertama yaitu pengujian data yang berasal dari situs berita Tribunnews.com sehingga diperoleh total 500 dokumen berita. Algoritma stemming yang digunakan pada skenario X1 adalah metode Nazief & Andriani, dan skenario X2 menggunakan metode Tala. . Tahap pertama adalah melakukan proses pra-pemrosesan teks dan dilanjutkan dengan proses stemming. Setelah pengujian skenario pertama dilanjutkan dengan pengujian skenario kedua yaitu Y1 dan Y2 yang sumber datanya berasal dari abstrak laporan skripsi mahasiswa. Pengujian Y1 menggunakan algoritma Nazief & Andriani, sedangkan pengujian Y2 menggunakan algoritma Tala.

Dari skenario yang dilakukan selanjutnya akan dilakukan pengukuran akurasi. Pengukuran akurasi dilakukan berdasarkan jumlah kata yang ditemukan memiliki jenis kata dasar dibagi dengan jumlah kata dalam satu dataset. Selain itu juga perhitungan berapa lama waktu yang dibutuhkan metode dalam proses stemming dengan sumber data yang disediakan. Perhitungan waktu dihitung dari selisih waktu setelah proses selesai dikurangi waktu dimulainya proses. Dari hasil tersebut akan diperoleh rata-rata setiap skenario untuk akurasi dan waktu pengerjaan. Setelah melakukan serangkaian uji stemming dengan kedua algoritma yang digunakan, diperoleh hasil seperti pada Tabel XIII dan XIV.

TAMPUXIII
DATA HASIL UJI X1

TIDAK	Kode	Total	BENAR	PALSU	BENAR %	PALSU %	Waktu
1	012	254	180	68	70,8	26,77	65,70
2	018	110	85	30	77,2	27,27	20,99
3	019	170	155	15	91,1	8,82	40,00
4	021	127	108	15	85,0	11,81	20,20
5	023	150	90	68	60	45,33	40,14
…	…	…	…	…	…	…	…

DOI : http://dx.doi.org/10.25139/inform.v8i1.5569

TIDAK	Kode	Total	BENAR	PALSU	BENAR %	PALSU %	Waktu
500	026	70	63	4	90	5.714	8.054
	Total	58.959	46,4	12,5	39, 6	12,02	58,95
	Rata-rata	115.430	91,7	24,7	78,5	21,53	23,75

TMAMPUXIV
DATA HASIL UJI X2

TIDAK	Kode	Total	BENAR	PALSU	BENAR %	PALSU %
1	300712	254	180	68	70,86	65,70
2	300718	110	85	30	77,27	20.993
3	300719	170	155	15	91,17	40,00
4	300721	127	108	15	85,03	20,20
5	300723	150	90	68	60	40,14
...	...	...	...	...	...	...
500	300726	70	63	4	90	8.054
	Total	59.652	34.232	20.331		5.022
	Rata-rata	117.230	86,56	35,24	64,37	10,75

Dari data proses uji coba yang tertera pada Tabel XIII dan Tabel XIV dapat disimpulkan bahwa untuk skenario pengujian 1 X1 dan X2 dengan data yang bersumber dari berita di situs Tribunnews.com, jumlah kata pada Tabel XIII sebanyak 59.959 kata, dengan rata-rata 115 katakan. Nilai rata-rata akurasi yang diperoleh adalah 78,47% dan rata-rata waktu yang dibutuhkan dalam proses adalah 23,75 detik, sedangkan untuk Tabel XIV jumlah kata yang diperoleh adalah 59.652 dengan rata-rata 117 kata. Untuk akurasi diperoleh nilai rata-rata sebesar 64,37% dan waktu rata-rata yang dibutuhkan sebesar 10,75 detik. Berdasarkan data tersebut, hasil pengujian keakuratan data sumber berita Tribunnews.com, algoritma Nazief & Adriani memiliki nilai lebih tinggi dibandingkan Tala yaitu 78,47%. Sebaliknya dari segi waktu pengerjaan, algoritma Tala lebih cepat dibandingkan menggunakan Nazief & Adriani yaitu 10,75 detik. Setelah dilakukan pengujian skenario X1 dan X2, selanjutnya dilakukan proses pengujian skenario Y1 dan Y2, dimana data pengujian ini berasal dari abstrak dokumen skripsi mahasiswa. Dari pengujian yang dilakukan, hasilnya terdapat pada Tabel XV dan Tabel XVI.

TMAMPUXV
DATA HASIL UJI Y1

TIDAK	Mengajukan Nama	Total	BENAR	PALSU	BENAR %	PALSU %
1	a.pdf	374	310	64	82,89	226, 7
2	b.pdf	432	402	30	93,06	308,8
3	c.pdf	156	140	16	89,74	68,0
4	d.pdf	175	153	22	87,43	75,6
5	e.pdf	142	122	20	85,92	104,1
...	. .	...	...	...	...	...
50	n.pdf	117	101	16	86,32	44.384
Total		7.438	6.114	1.211		3.407,9
Rata-rata		148	124,68	26,28	82,63	65,2

TMAMPUXVI
DATA HASIL UJI Y2

TIDAK	Mengajukan Nama	Total	BENAR	PALSU	BENAR %	PALSU %
1	a.pdf	190	144	46	75,8	53,6
2	b.pdf	238	179	59	75,2	6,8
3	c.pdf	113	90	23	79,6	15,9
4	d.pdf	178	131	47	73,6	28,5
5	e.pdf	142	102	40	71,8	33,3
...	...	...	...	...	...	...
50	n.pdf	127	90	37	70,9	27,9
	Total	7.067	5.013	2.053		1.705
	Rata-rata	140	99.420	40.100	68,6	32,19

Dari hasil pengujian yang disajikan pada Tabel XV dan XVI dapat dijelaskan bahwa Tabel XV menggunakan algoritma Nazief & Andriani sebagai metode stemming dan ditemukan 7.438 kata dengan rata-rata 148 kata. Nilai rata-rata akurasi yang diperoleh sebesar 82,63% dengan rata-rata waktu yang dibutuhkan 65,2 detik, sedangkan untuk Tabel XVI dengan menggunakan algoritma Tala ditemukan 7.067 kata dengan rata-rata 140 kata. Nilai rata-rata diperoleh untuk akurasi sebesar 68,6% dengan waktu rata-rata 34,19 detik. Berdasarkan data Tabel XV dan XVI di atas dapat disimpulkan bahwa akurasi algoritma dari Nazief & Andriani mempunyai nilai yang lebih tinggi dibandingkan Tala yaitu sebesar 82,63%. Sebaliknya dari segi waktu yang dibutuhkan, algoritma Tala mempunyai waktu yang lebih cepat dibandingkan dengan algoritma Nazief & Andriani. Adriani. yaitu 32,19 detik.

IV. KESIMPULAN

Nilai rata-rata akurasi yang diperoleh dengan menggunakan dataset baru 500 berita metode Nazief & Andriani adalah 78,47% yang berjumlah 58.964 kata. Rata-rata akurasi metode Tala sebesar 65,29% dengan jumlah kata sebanyak 59.652 kata. Nilai rata-rata akurasi yang diperoleh dengan menggunakan dataset dokumen 50 laporan skripsi metode Nazief & Andriani adalah 82,63% % dengan jumlah kata sebanyak 7.438 kata. Sebaliknya, metode Tala memiliki nilai akurasi sebesar 68,6% yang berjumlah 7.067 kata. Nilai rata-rata waktu proses tercepat menggunakan dataset baru metode Nazief & Andriani sebesar 25,27 detik, dan metode Tala sebesar 11,08 detik. Untuk dataset dokumen rata-rata waktu pengerjaannya paling cepat dengan metode Nazief & Adriani 65,2 detik dan metode Tala 32,19 detik.

Dari segi keakuratan proses stemming, algoritma Nazief & Andriani mempunyai akurasi yang lebih baik dibandingkan dengan algoritma Tala. Sedangkan dari segi kecepatan waktu pemrosesan, algoritma Tala memiliki kecepatan yang lebih baik dibandingkan dengan algoritma Nazief & Andriani.

Pekerjaan lebih lanjut sedang dikembangkan untuk lebih memahami keandalan algoritma stemming dalam hal akurasi dan kecepatan proses stemming. Kumpulan data perlu direproduksi tidak hanya dari dua sumber tetapi dari setidaknya tiga sumber data. Dibutuhkan banyak sumber data agar hasil pengujian bisa lebih optimal.

RREFERENSI

[1] Afuan L, "Membendung Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Porter".*Jurnal Telematika*Jil. 6 No.2, hal.34-40, 2013.
[2] Agusta L. "Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia". *Konferensi Nasional Sistem dan Informatika 2009*, November 2009.
[3] Novitasari, D., "Perbandingan Algoritma Stemming Porter Dengan Arifin Setiono Untuk Menentukan Tingkat Akurasi Kata Dasar", *Jurnal String*, Jil. 1 No.2, hal.120-129, 2016.
[4] Nopiyanti D., Sekarwati, KA, "Aplikasi Pencarian Kata Dasar Dokumen Berbahasa Indonesia Menggunakan Metode Porter Stemming Menggunakan PHP & MYSQL",*Prosiding Ilmiah Nasional*

DOI : http://dx.doi.org/10.25139/inform.v8i1.5569

Seminar Komputer dan Sistem Intelijen (KOMMIT 2014). Oktober 2014.
[5] Utomo, MS, "Implementasi Tala Stemmer pada Aplikasi Berbasis Web".*Jurnal Teknologi Informasi DINAMIS,*Jil. 18, No.1.Hal. 41-45, ISSN : 0854-9524, 2013.
[6] Wiguna, PB,S., Hantono,BS, "Penyempurnaan Algoritma Porter Stemmer Indonesia Berbasis Metode Morfologi dengan Menerapkan 2 Level Morfologi serta Aturan Kombinasi Prefix dan Suffix",*JNTETI,* Jil. 2 No.2, hlm.1-6, ISSN : 2301 – 4156, 2013.
[7] Indriyono, BV, Utami E, Sunyoto, A. "Pemanfaatan Algoritma Porter Stemmer untuk Proses Klasifikasi Jenis Buku Berbahasa Indonesia Dalam",*Jurnal Informatika Buana,*Jil. 6 No.4, hal.301-310, 2015.
[8] Ariyani, PF, , Rahmala A., Juliasari, N." Implementasi Metode Tala Stemming dan Fungsi Jaccard Pada Aplikasi Katalog Perpustakaan",*Seminar Nasional Inovasi dan Penerapan Teknologi di Industri 2019,*Februari 2019.
[9] Ghazvini, A., dkk, "Algoritme stemming untuk tenses yang berbeda untuk meningkatkan kamus bahasa Persia", 2012.*Simposium IEEE tentang Elektronika Industri dan Aplikasi,*September 2012.
[10] Pramudita, HR, "Implementasi Algoritma Stemming Nazief & Adriani Dan Persamaannya Pada Penerimaan Judul Skripsi",*Jurnal Ilmiah DASI*, Jil. 15 No.04, hlm.15-19, ISSN : 1411-3201
[11] Yulianto, MA,, Nurhasanah, "Pengaruh Stemming Nazief & Adriani terhadap Kinerja Algoritma Rabin-Karp dalam Mendeteksi Kesamaan Teks",*Jurnal Informatika Universitas Pamulang,*Jil. 6, No.4, hlm.880-886, ISSN : 2541-1004, 2021.
[12] Sugiyono,*Metode Penelitian Kuantitatif, Kualitatif, dan R&D.* Bandung : Alfabeta. 2017
[13] Hasibuan, Z. Metodologi Penelitian Bidang Komputer dan Teknologi Informasi. Jakarta : Universitas Indonesia.
[14] Arikunto, S.*Prosedur penelitian.* Jakarta: Rineka Cipta.
[15] Parwita, WGS, "Menguji Akurasi Sistem Rekomendasi Penyaringan Berbasis Konten",*Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer,*Jil. 14, No.1, hlm.27-32, ISSN : 1858-4853, 2019.
[16] Prihatini, PM, dkk, "Algoritma Stemming untuk Pengolahan Teks Berita Digital Indonesia",*Jurnal Internasional Teknik dan Teknologi Berkembang,*Jil. 2, No.2, hlm.1-7, ISSN : 2579-5988, 2017.
[17] Saifudin, A., Verdaningroem, NJ,M.," Penerapan Kamus Dasar Algoritma Porter untuk Mengurangi Kesalahan Stemming Bahasa Indonesia",*Jurnal Teknologi,*Jil. 10, No.2, hal.103-112. ISSN : 2085 – 1669, 2018.

DOI : http://dx.doi.org/10.25139/inform.v8i1.5569